


In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing

Mehrgol Tiv , Elisabeth O'Regan and Debra Titone

McGill University, Montréal, Canada

Research Article

All anonymized materials are available on the Open Science Framework: <https://osf.io/m2h93/>

Cite this article: Tiv M, O'Regan E, Titone D (2021). In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing. *Bilingualism: Language and Cognition* **24**, 918–931. <https://doi.org/10.1017/S1366728921000225>

Received: 15 July 2020
Revised: 29 March 2021
Accepted: 7 April 2021
First published online: 2 June 2021

Keywords:

mentalizing; bilingualism; inferences; theory of mind; social cognition; pragmatics; language diversity

Address for correspondence:

Mehrgol Tiv or Debra Titone, Ph.D.
Department of Psychology,
McGill University 2001 McGill College Ave.
Montreal, Quebec, H3A 1G1, Canada
Telephone: (514) 398-1778 / Fax: (514) 398-4896
E-mail: mehrgol.tiv@mail.mcgill.ca;
debra.titone@mcgill.ca

Abstract

Mentalizing, a dynamic form of social cognition, is strengthened by language experience. Past research has found that bilingual children and adults outperform monolinguals on mentalizing tasks. However, bilingual experiences are multidimensional and diverse, and it is unclear how continuous individual differences in bilingual language experience relate to mentalizing. Here, we examine whether individual differences in bilingual language diversity, measured through language entropy, continuously pattern with mentalizing judgments among bilingual adults, and whether this relationship is constrained by first vs. second language reading. We tested sixty-one bilingual adults on a reading and inference task that compared mental state and logical inferences. We found that greater language diversity patterned with higher mentalizing judgments of mental state inferences across all readers, and that L2 readers attributed more mentalizing to logical inferences compared to L1 readers. Together, we found evidence of a positive relationship between continuous individual differences in bilingual language diversity and mentalizing.

Introduction

Mentalizing is a dynamic and flexible form of social cognition that aids in the understanding of others' intentions and actions. Its dynamic nature arises from nuances in situational context or life experience that alter mentalizing in some way, even during adulthood (e.g., Dumontheil, Apperly & Blakemore, 2010; Valle, Massaro, Castelli, Sangiuliano Intra, Lombardi, Bracaglia, & Marchetti, 2016). In some situations, such as negative interactions with out-group members, mentalizing may be de-emphasized or withheld (leading to dehumanization). Conversely, mentalizing may be strengthened through other experiences and individual differences that focus on understanding others (Baimel, Birch & Norenzayan, 2018; Conway, Coll, Cuve, Koletsi, Bronitt, Catmur & Bird, 2019; Harris, 2017; Kidd & Castano, 2013; Slaughter & Repacholi, 2003). Across the lifespan, greater language proficiency boosts mentalizing performance (Milligan, Astington & Dack, 2007; Pyers & Senghas, 2009; Warnell & Redcay, 2019), and some work has linked bilingualism, as a categorical variable, to stronger mentalizing skills (e.g., Antoniou, 2019; Goetz, 2003, Kovács, 2009). In this paper, we further investigate the relationship between bilingualism and mentalizing, by examining how a CONTINUOUS individual difference in bilingual language experience – specifically, language diversity – relates to mentalizing among bilingual adults.

Mentalizing, which contributes to Theory of Mind, manifests through inferences or deductions about the thoughts, feelings, beliefs, and goals of another person (Frith & Frith, 1999; Harris, 2017). Importantly, it plays a role in how inferences are shaped, and how predictions of others' behaviours occur (Apperly, 2008; Premack & Woodruff, 1978). Accordingly, mentalizing helps us answer the question of why someone may do or say something (or not do or say something). As a result, mentalizing is directly implicated in many real-world scenarios including the detection and use of sarcasm and irony in everyday conversation (e.g., Antoniou, 2019; Sperber & Wilson, 2002; Tiv, Rouillard, Vingron, Wiebe & Titone, 2019; Tiv, Deodato, Rouillard, Wiebe & Titone, 2020).

Among adults, past work on mentalizing has largely relied on a limited number of tasks or measures. These have included questionnaires (e.g., Baron-Cohen & Wheelwright, 2004; Davis, 1980; Jolliffe & Farrington, 2006), false belief tasks (Sally and Anne task; Baron-Cohen et al., 1985), false belief stories (e.g., Conway et al., 2019; Fletcher, Happé, Frith, Baker, Dolan, Frackowiak and Frith, 1995; Kanske, Böckler, Trautwein, Parianen Lesemann & Singer, 2016; Pino & Mazza, 2016; Saxe & Kanwisher, 2003), Reading the Mind in the Eyes task (Baron-Cohen, Wheelwright, Hill, Raste & Plumb, 2001), the Director Task (Dumontheil et al., 2010), Attribution of Intentions task (Sarfati, Hardy-Baylé, Besche & Widlöcher, 1997), the Multifaceted Empathy Test (Dziobek, Rogers, Fleck, Bahnemann, Heekeren, Wolf & Convit, 2008), and others (e.g., Todd, Simpson & Tamir, 2016). Although the majority of

these studies have used linguistic stimuli to assess mentalizing (e.g., stories, conversations, essays), few have addressed mentalizing from a linguistic perspective (cf. Ferstl & von Cramon, 2002). Indeed, mental state inferences are “first and foremost inferences” (Harris, 2017) – meaning that they are conclusions people reach by integrating prior evidence, which often take shape as linguistic text or speech.

Linguistic inferences

Inference-making occurs automatically at all levels of language processing. Inferencing is particularly crucial during text comprehension as readers construct a situation model, or mental representation, of what the text is about (Graesser, Singer & Trabasso, 1994). Locally, inferences contribute to the coherence between words, aiding in pronoun-antecedent bridging and case role assignment of nouns. Beyond building local coherence, inferences can be elaborative and relate to the global structure of the text, thereby contributing to meaning that is pragmatically “between the lines” (Graesser et al., 1994; reviewed in Snow, 2002). These pragmatic inferences typically occur across sentences and involve using prior knowledge to fill in missing information that is implied by the text, often to establish explanations for behaviors or actions (Graesser et al., 1994; Harris & Monaco, 1978; Johnson-Laird, 1993; Kispal, 2008). Indeed, prior knowledge for coherence may be based in LOGICAL, non-social aspects of the context, including general world knowledge of causality (e.g., if X then Y). However, coherence may also be achieved by drawing upon MENTALIZING or privileged information in an actor’s mind that is subsequently used to explain her observable behaviours in reality (e.g., she did X because Y) (Astington & Gopnik, 1988; Gaudreau et al., 2015). Distinguishing these two forms of pragmatic inferences (logical vs. mental state) is critical in understanding the unique contributions of mentalizing, as opposed to non-mental-state coherence building during text comprehension.

Pragmatic inferences can take many forms during language processing, particularly when an implicit meaning, or implicature, is intended by a speaker or writer (Antoniou, 2019; Sperber & Wilson, 2002). Interestingly, greater second language experience or proficiency have been reported to enhance implicature performance (reviewed in Antoniou, 2019), particularly in the form of metaphorical and ironic language (e.g., Johnson & Rosano, 1993; Tiv et al., 2019b; 2020). For example, past work from our group examined the relationship between continuous bilingual language experience and verbal irony, which is a common form of implicature that depends on mentalizing for successful comprehension (Banasik, 2013; Filippova & Astington, 2010; Sperber & Wilson, 2002). Specifically, Tiv et al. (2019b) found that greater second language proficiency patterned with greater self-perceptions of general sarcasm use (across all languages known). Tiv et al. (2020) found a similar link between greater second language proficiency and more sensible judgments and faster response times to non-canonical irony forms in the first language. These are among the few studies that have examined the role of CONTINUOUS, INDIVIDUAL DIFFERENCES among bilinguals on processes that involve the building blocks of mentalizing.

Mentalizing and bilingualism

There is a growing literature demonstrating that bilinguals categorically exhibit greater mentalizing capacities than monolinguals across the lifespan. First, among children, bilinguals

consistently outperform monolinguals in the classic Sally-Anne false belief task across many language pairs and regions (e.g., Goetz, 2003; Kovács, 2009; Rubio-Fernández & Glucksberg, 2012), which was corroborated by a meta-analysis that controlled for differences in language proficiency between the groups (Schroeder, 2018). Next, among adults, bilinguals also exhibit less egocentric bias than monolinguals in a visual world adaptation of the Sally-Anne false belief task (Rubio-Fernández & Glucksberg, 2012), in a spatial perspective-taking task (Navarro & Conway, 2020), and within academic writing (Hsin & Snow, 2017). Lastly, among older adults, bilinguals in Singapore (i.e., a highly multilingual context) maintain accuracy and efficiency when forming pragmatic inferences, but monolinguals in the United Kingdom, where monolingualism is more prevalent, demonstrate an age-related decline in making pragmatic inferences (Sundaray, Marinis & Bose, 2018; cf. Antoniou & Katsos, 2017).

While the reported links between bilingualism and mentalizing are compelling, there are two discernible limitations of past work. First, many have evaluated mentalizing through the traditional false belief paradigm (e.g., Sally and Anne Task), which may not capture all aspects of mentalizing (e.g., Bloom & German, 2000). Second, and more critically, these studies generally compare bilinguals to monolinguals monolithically, which undercuts the vast diversity in language experience found within each group and contributes to the dominant ideology of monolingualism as a gold standard (Baum & Titone, 2014; Gullifer & Titone, 2019, 2020), or monolingual hegemony (e.g., Ortega, 2018). Indeed, bilinguals (and monolinguals) vary in meaningful and consequential ways within group. For example, consider two proficient English–French bilinguals: Fatima uses English for most of her everyday life but speaks in French with her grandparents on the phone. Maryam uses both English and French across all her social circles. A typical group comparison would categorize these two experiences as “bilingual”, but it is plausible that Maryam’s consistent recruitment of multiple languages has cultivated greater attention to environmental social cues in predicting the linguistic preferences of her conversational partners, which over time may exercise her mentalizing capacity. To mitigate the erasure of these and other important individual differences, more and more studies are relying on the continuous, rather than categorical, assessment of bilingual language experiences (discussed in Gullifer & Titone, 2020).

One relatively new, continuous measure is language entropy, introduced by Gullifer and Titone (2019), to characterize language diversity. This characterization applies Shannon’s Entropy from Information Theory, a measure of diversity or overall uncertainty in a system, to quantify the distribution of usage across all known languages (also used in Bice & Kroll, 2019; De Bruin, 2019). We can apply this principle to the previous example of Fatima and Maryam: Maryam: Maryam, who uses English and French in a fully balanced manner (50–50), has high entropy, meaning that at any point in time she has an equally likely chance of using any of her two languages (maximal uncertainty). In contrast, Fatima, who is functionally monolingual in English, has low entropy, meaning that there is more certainty in her language use. Language entropy captures the overall balance of multilingualism more efficiently than proportion of usage, and it is robust to cases where more than two languages are used. For example, now imagine that Fatima and Maryam each grew up with three languages but now do not use their first language (L1). On the surface, both of their proportion of L1 use would be 0; however, Fatima may actively use her second (L2) and third (L3) languages (high multilingualism), whereas Maryam only actively uses her L2

(functional monolingualism). These nuances would be reflected through language entropy, but not in proportion of L1 usage or other language-specific measures.

Thus, language entropy can quantify the degree of bilingualism, in terms of people's experiences navigating dual-language social environments. Gullifer and Titone (2020) found that greater environmental linguistic diversity, as measured through entropy, predicts stronger proactive executive control in particular, suggesting that this index relates to attendance of contextual information, including social context. Similarly, Fan, Liberman, Keysar, and Kinzler (2015) identified that diverse sociolinguistic environments enhance mentalizing capacities, even among monolinguals, implicating the social consequences of bilingualism on flexible social cognition (the SOCIAL-PRAGMATIC ACCOUNT OF bilingual social cognition; see also Tiv et al., 2019b, 2020). Collectively, these findings provide further evidence that greater bilingual experience may globally (i.e., not specific to operating in one language) promote pragmatic awareness and attention to social information – core components of mentalizing. Thus, we hypothesize that through this enhanced social-pragmatic flexibility, greater language diversity will relate to stronger mentalizing skills among bilingual adults.

Despite expected global facilitation of bilingualism on mentalizing, we return to Harris' (2017) point that mental state inferences are “first and foremost inferences,” meaning that they are influenced by low-level linguistic processing and language proficiency. Some research has revealed that low L2 fluency relates to greater difficulty in making “complex inferences” (i.e., pragmatic inferences) during text reading, whereas high proficiency L2 readers are more likely to draw these inferences (Horiba, 1996; Rai, Loschky, Harris, Peck & Cook, 2011). Others have found that readers construct stronger situation models during L1 reading compared to L2 reading (Zwaan & Brown, 1996). In contrast, some research has shown that the social benefits of bilingualism during language processing may compensate for or outweigh language-specific challenges (e.g., Antoniou, 2019; Ramírez-Esparza, García-Sierra & Jiang, 2020; Verhagen, Grassmann & Küntay, 2017). For example, Foucart, García, Ayguasanosa, Thierry, Martin, and Costa (2015) found that, during online sentence reading, late L2 readers relied more on speaker identity during text comprehension than L1 readers, suggesting a heightened sensitivity to social, pragmatic, and contextual information. Thus, it is possible that the relationship between bilingual language experience and mentalizing may be constrained by whether the reader is processing in their first or second language, even if all readers are collectively bilingual. The consideration of L1 vs. L2 reading should be underscored in task designs that are inherently linguistic (i.e., rely on language processing, such as text comprehension), such as the present work (see also Tiv et al., 2019a for L1/L2 differences in figurative language reading).

Present study

We used a two-sentence vignette reading and judgment task to assess the extent to which logical, mental state, and incoherent (control) inferences were rated as coherent and reliant on mentalizing. For example, if a character was said to “lock the front door”, the preceding context sentence may have been that she “took out the house keys” (logical), “read about the increase in crime” (mental state), or “had a fancy pencil case” (incoherent).

In this paper, we pose three key questions. First, do individual differences among bilingual readers in language diversity, as

reflected by language entropy, relate to their mentalizing ratings? Here, we compare coherence and mentalizing ratings for logical vs. mental state inferences. Given past findings on the social consequences of bilingualism on mentalizing (e.g., Fan et al., 2015), we predict that greater language diversity will pattern with higher mentalizing and coherence ratings to mental state inferences, in particular. Second, does the relationship between language diversity and mentalizing ratings vary as a function of L1 vs. L2 reading? Based on past evidence of L2 sensitivity to social information, we expect that L1 readers may surpass L2 readers in detecting coherence among logical inferences, but that L2 readers' reliance on social information will facilitate their detection (i.e., higher ratings) of coherence and mentalizing among mental state inferences. Third, are there observable differences in global reading time when readers encounter logical vs. mental state inferences in a short vignette context? We hypothesize that reading times for mental state inferences may be longer than logical inferences, due to processing costs, for all readers, but that this relationship may be mediated by individual differences in bilingual language experience.

Methods

Participants

Sixty-six healthy bilingual adults were recruited from, and lived, in the linguistically diverse city of Montréal, which is situated in Québec, Canada. Québec's provincial government is officially French speaking (i.e., French is the legal language). At the national level, Canada has two official languages: English and French. As a result of this – among other factors, such as proximity to the northern states of the U.S.A. where English is a dominant language, as well as Indigenous nationhood, globalization, urbanization, and immigration – many languages can be overheard in Montréal. Nevertheless, legal mandate from the province directs that all businesses, street signs, and advertisements predominantly feature French. Taken altogether, most inhabitants of the city experience some degree of exposure to multiple languages, even if they themselves do not use these languages to communicate within their network.

We recruited participants using flyers, online advertisements, and word of mouth. Recruitment materials indicated that eligible participants would be comfortable reading sentences in English and French (although all materials were in English). Following data preprocessing, we discarded five individuals who demonstrated excessively long or short response times (see *Data Preprocessing*). This left us with sixty-one bilingual adults aged 18–35 (mean age = 21.7 years). Of this sample, 47 individuals selected female as their gender, ten male, and four queer or non-binary from the options provided (see Table 1). The racial-ethnic composition of the sample was predominantly White (40). Of the remainder, seven selected East Asian, three Black, two Middle Eastern, one Southeast Asian, and seven Multiethnic (one individual preferred not to respond) from the options provided (see Table 1). The majority of participants were enrolled in an undergraduate program (51), three had completed a high school or college education, and seven were in or completed their graduate studies.

We operationally defined language group (L1 vs. L2) by age of English acquisition. We categorized anyone who was exposed to English in their first year of life as in the L1 English group, and anyone who was exposed to English after their first year of life as the L2 English group. Individuals who were exposed to

Table 1. Descriptive statistics of 61 participants split across the L1 English and L2 English language groups.

	English L1 (N = 31)			English L2 (N = 30)		
	Mean	SD	Range	Mean	SD	Range
Age	21.77	3.04	18 to 31	21.61	2.31	18 to 28
Age of Acquiring English	0	0	0 to 0	7.63	3.14	1 to 14
Age of Acquiring non-English language	4.97	3.52	0 to 19	0	0	0 to 0
Daily Percent English Use	83.08	13.34	50 to 97	53.34	22.38	5 to 100
Daily Percent non-English Use	16.92	13.34	3 to 50	46.66	22.38	0 to 95
Language Diversity (entropy)	0.59	0.27	0.19 to 1.01	0.91	0.31	0 to 1.58
Parental/Guardian Socioeconomic Status ^a	5.22	1.04	2.5 to 7	5.28	1.23	2 to 7
	Count	Percent of Subsample		Count	Percent of Subsample	
Gender^b						
Female	27	87.10%		20	66.67%	
Male	3	9.68%		7	23.33%	
Queer/Non-binary/multiple selections	2	6.45%		3	10.00%	
Race/Ethnicity^c						
Black	1	3.23%		2	6.67%	
East Asian	3	9.68%		4	13.33%	
Middle Eastern	0	0.00%		2	6.67%	
Southeast Asian	0	0.00%		1	3.33%	
White	22	70.97%		18	60.00%	
Multiracial	4	12.90%		3	10.00%	
Prefer not to answer	1	3.23%		0	0.00%	
Education^d						
Graduate School (PhD/MD/JD)	3	9.68%		0	0.00%	
Graduate School (Masters)	2	6.45%		2	6.67%	
University/College	26	83.87%		25	83.33%	
CEGEP/Associate's Degree	0	0.00%		2	6.67%	
Secondary/High School	0	0.00%		1	3.33%	
Place of birth						
In Canada	27	87.10%		12	40.00%	
Outside of Canada	4	12.90%		18	60.00%	

^aParental/guardian socioeconomic status was calculated by converting each of the two parent/guardians' highest education level into an ordered, numerical value (1–7) and averaging across the two. In cases of single parenthood/guardianship, this value reflects the single education level.

^bParticipants selected all the gender options that best represented them from the following list: female, male, trans, intersex, queer/non-binary, and other.

^cParticipants selected all the racial and ethnic options that best represented them from the following list: Black, White, East Asian, Southeast Asian, Latin American, Middle Eastern, Indigenous, Pacific Islander, Other, Prefer not to answer. Anyone who selected more than one option is represented as Multiracial in this table.

^dEducation refers to the highest degree obtained by the participant.

English (the language of the task) and another language in the first year of life (simultaneous bilinguals) were excluded from analysis. The L1 English group comprised thirty-one participants who also acquired knowledge of one or more additional languages later in life. The thirty L2 English group participants reported being exposed to a language other than English in their first year of life and acquiring English (and sometimes additional languages) later in life. Among them, twenty-seven had French as their first language, one had both French and Arabic as the first languages, and two others were exposed to Mandarin or Bengali in their first year of life. All participants self-reported a

comfortable working proficiency of English. In total, twenty-eight participants reported knowing two languages, twenty-two reported knowing three languages, and eleven reported knowing four languages. Additional information about the two language groups can be found in [Table 1](#).

Materials

The text materials were originally created for this experiment and future eye tracking follow-up studies, though some were slightly adapted from past work (e.g., Ferstl & von Cramon, 2002;

Table 2. Two example items in the three inference type conditions (logical, mental state, and incoherent). The first sentence (context) varies across inference type conditions, but the second sentence (action) is identical across inference type conditions.

		Logical Inference	Mental State Inference	Incoherent
Example 1	Sentence 1 (Context)	Jane took out the house keys.	Jane read about the increase in crime.	Jane had a fancy pencil case.
	Sentence 2 (Action)	She locked the front door that day.	She locked the front door that day.	She locked the front door that day.
Example 2	Sentence 1 (Context)	Mark turned down the music at home.	Mark received the bad news at home.	Mark was too short to reach the shelf.
	Sentence 2 (Action)	His apartment got much quieter that day.	His apartment got much quieter that day.	His apartment got much quieter that day.

Nadig & Ozonoff, 2007; Lavoie, Vistoli, Sutliff, Jackson & Achim, 2016). The texts consisted of 138 sentence-pair item sets, in three inference type conditions (see Table 2 for examples). Each item was composed of two sentences: the context and the action. The first was a context sentence, which was unique across the three condition types and described a situation involving a character. In contrast, the second was an action sentence, which was identical across the three condition types and described an action that was either somehow related to the first sentence or unrelated to the context. This meant that the second sentence of all three conditions began with a gender consistent subject pronoun (80% of items) or possessive pronoun (20% of items) referring back to the character in the context sentence (The gender of the character was controlled within items to account for any potential intergroup effects, and the perceived gender of each character was measured at the end of the task.). By constructing the items in this manner, we ensured that any differences in reading behavior of the target action sentence in future eye tracking work would be conducted on a set of literally identical words whose pragmatic meanings change as a function of the preceding, non-target context sentence. This design gave rise to 552 unique sentences (414 unique contexts + 138 unique actions) with an average of 13 words across both sentences – available on the Open Science Framework: <https://osf.io/m2h93/>

The critical difference across the three conditions was the type of inference required to connect the context and action sentences. In the logical condition, the context and action sentences could be connected through a non-social logical, causal, or deductive inference. Some of these inferences involved physical cause/effects, a logical action sequence, or some type of non-social bodily response like shivering from being in the cold (see Van Overwalle, 2011 for a discussion of how these physical inferences about people do not engage mentalizing brain regions). For example, in “Jane took out the house keys. She locked the front door that day”, the logical inference may be that Jane used the house keys to lock the front door. Indeed, given that the logical inference condition, just as the other two conditions, featured the actions of a human character, it is possible that the reader experienced some degree of mentalizing. However, the logical condition was designed to rely substantially less on mentalizing than the mental state condition, which we affirmatively tested as a manipulation check.

In the mental state condition, the context and action sentences were designed to be connected through an inference involving an understanding of the character’s mental state, emotions, intentions, goals, and beliefs. Critically, the action was identical to the logical inference, but the context differed, which we expected

would give rise to an inference that relied more on how the character is feeling or thinking. For example, in “Jane read about the increase in crime. She locked the front door that day” one might infer that Jane’s mind was preoccupied with nervous thoughts of someone breaking into her home, which motivated her action to lock the front door. In designing these inferences, we strove to capture a wide range of emotional and cognitive experiences, from fear and sadness (Table 2) to joy and excitement. Whereas in some items there was explicit or implicit mention of a second character (e.g., a potential burglar in this example), this was not the case across all items (e.g., Example 2 in Table 2).

The third condition served as a baseline control for situations in which no explanatory inference could have easily connected the context and action sentences (excluding the pronoun-antecedent bridging inference to the character’s name). In this incoherent condition, the context was created to be disconnected from the action, which was still identical to the logical and mental state conditions. For example, in “Jane had a fancy pencil case. She locked the front door that day” no clear explanatory inference would aid in connecting the relationship between the pencil case and locking the front door.

Given that a future goal was to track on-line reading of these items with eye tracking, we added additional words that did not change the meaning of the sentence to the end of the first sentence, as a spillover processing region (e.g., Poynor & Morris, 2003). These words all conveyed generic, temporal information such as “that day” or “in that instant.” Additionally, in some items, we used a definite article (e.g., “the front door”) where a possessive article may have sounded more natural (e.g., “her front door”). We did this for two reasons: first, to avoid additional pronoun-antecedent bridging demands on the reader, and second, to be consistent about the amount of gender-revealing information in the item in order to avoid inter/intragroup mentalizing differences (e.g., Todd, Hanko, Galinsky & Mussweiler, 2011; Todd et al., 2016).

Language History Questionnaire

All participants completed a short language history questionnaire created by our group, which probed various aspects of their language experience and demographic background. For example, the questionnaire assessed at what age each language was first acquired, what percent of each day each language was used, and whether each language was still being used. The responses to this questionnaire were used to divide the sample into L1 vs. L2 readers and compute language entropy (see *Computing Language Diversity through Language Entropy*).

Procedures

All participants provided written consent prior to participating in the experiment. The materials and procedures in this experiment were approved by the McGill Research Ethics Board. First, participants completed the inference reading task, which involved reading the text materials described above and making two judgments about each item. Prior to the start of this task, all participants completed six practice trials and were ensured by the researcher to have understood the instructions. In this task, participants were randomly presented with an item in one of the three conditions, which appeared in full on the computer screen. Participants were instructed to silently read for comprehension and press the spacebar when they had done so. There was no upper time limit for global reading time, which coarsely measured comprehension time for the entire item. Following the press of the spacebar, the first judgment question on linguistic coherence between the context and action sentences appeared below the item which remained on the screen. Here, participants were asked to rate whether there was a relationship between the sentences, from 1 (not at all) to 5 (completely). Participants indicated their choice by pressing a number on the keyboard, and we measured their response time to this probe (see Supplementary Materials for all reaction time results). From there, a second judgment question replaced the first question on the screen, while the item still remained on the screen. Here, participants were asked to rate to what extent the relationship between the sentences relied on mentalizing. Mentalizing was defined to all participants at the start of the task as meaning consideration of the thoughts, beliefs, emotions, and goals of the story character. Again, participants rated the need for mentalizing on a scale of 1 (not at all) to 5 (completely) by pressing a number on the keyboard while the response time was measured. At that point, participants moved on to the next item and repeated this exercise for all 138 items. Each item only appeared in one of the three inference type conditions, and the presentation order of items was randomly shuffled for each participant.

Computing language diversity through language entropy

Through the language history questionnaire, participants listed all the languages that they knew. Then, they indicated approximately for what percent of all daily conversations they used each of their known languages (all languages totalling 100%). For example, a participant may have indicated that they know English, French, and Farsi and use these languages in the following manner: 90% English, 5% French, and 5% Farsi. We calculated general language entropy, or language diversity, from the self-reported responses to this question using the languageEntropy package in R (Gullifer & Titone, 2018). This calculation is based on the following equation:

$$H = - \sum_{i=1}^n P_i \log_2(P_i).$$

Here, n represents the total possible languages ($n = 3$ for English, French, Farsi), and P_i represents the proportion that language _{i} is used ($P_{\text{English}} = 0.9$, $P_{\text{French}} = 0.05$, $P_{\text{Farsi}} = 0.05$). The total sum is multiplied by -1 to render a positive entropy value. The language entropy score of this particular example would be 0.569. The lower bound of entropy is zero, indicating that one language is used 100% of the time (functional monolingual). When the total number of possible languages is two, the upper bound of entropy is one, indicating maximal language entropy (50% one language, 50% other language). As the total possible number of languages increases (e.g., the English, French, Farsi trilingual described above), the upper bound of entropy also slightly increases.

Results

Data preprocessing and analysis

Prior to data analysis, we undertook a series of item-level, subject-level, and trial-level preprocessing steps. These steps yielded the removal of three out of 138 total full item sets (resulting in 135 analyzed items), one full participant, four participants' reaction times, and single trials that were slower than 10 seconds or faster than 500 ms. This left us with 61 participants (31 English L1, 30 English L2) for the judgment responses, and 57 (29 English L1, 28 English L2) for global reading time and the judgment reaction times. We considered the possibility that the methodological choice of having participants find the corresponding number key for their ratings may have resulted in noisy reaction time data. Additional noise may have been added to the reaction times because participants were asked to make multiple, subsequent judgments to the same item probe and may have been thinking about one rating as they reacted to the other one. For these reasons, we reported the results of the reaction time data in the Supplementary Materials, along with more specific details on the preprocessing steps.

All data were analyzed with linear mixed-effects regression models using the *lme4* package in R (Bates, Maechler, Bolker & Walker, 2014). Prior to examining the role of individual differences in bilingual language experience on mentalizing, we conducted a manipulation check of our items. To do this, we ran three maximal random effects (by-item and by-subject intercepts and slopes) models with inference type (helmert coded) as the sole predictor of coherence rating, mentalizing rating, and global reading time across all participants (Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017). To account for length differences across items, global reading time was standardized in two ways. First, we simply divided total reading time (ms) by the total number of characters for each item. Alternatively, we included total number of characters as a random slope by-subject. Since no differences were detected between these approaches, we report the first method in the manuscript and provide the results of the second method in Supplementary Materials. In the event that a model failed to converge, we dropped random slopes, as outlined by Barr, Levy, Scheepers, and Tily (2013), until convergence was achieved. Following the manipulation check, we will only report higher order interactions with individual difference variables to avoid repetition.

Following the manipulation check, we constructed a multi-level model that was tested separately for each dependent variable of interest (coherence rating, mentalizing rating, and global reading time. For coherence and mentalizing reaction times, see Supplementary Materials). This model coded for a three-way interaction between inference type (3 levels: logical, mental state, and incoherent), language diversity (scaled), and language group (2 levels: L1 vs. L2 English), and it included daily percent English use (the language of the task, scaled) and trial order (scaled) as continuous covariates. For the two categorical variables, language group was treatment coded with L1 English as the baseline and the three-level inference type variable was Helmert coded, such that the first contrast reflected the difference between logical vs. mental state inferences (C1-Log/Men) and the second contrast reflected the difference between the mean of logical and mental state inferences (coherent inferences) vs. incoherent (C2-Coh/Inc). This approach allowed us to assess subtle differences between logical and mental state inferences, which was a primary goal of this work, and also examine general

differences between sentence pairs that were created to be coherent (logical or mental state inferences) from those that were not (incoherent).

In these individual difference models, random effects structures were again built for both subjects and items so that we could generalize beyond our sample and item set. However, given that our primary aim was to affirmatively test for the role of specific subject-level individual differences in language diversity, we calculated all individual difference models with random intercepts-only. Full model outputs are available in Supplementary Materials, and analyses are available on the Open Science Framework: <https://osf.io/m2h93/>

Manipulation check

The linear mixed-effects regression assessing coherence ratings detected a significant main effect of condition at both contrasts (C1-Log/Men: $\beta = -0.239$, $SE = 0.032$, $t = -7.402$, $p < 0.001$; C2-Coh/Inc: $\beta = -1.024$, $SE = 0.0227$, $t = -45.173$, $p < 0.001$), indicating that incoherent items were perceived as less coherent than all coherent items and mental state items were perceived as less coherent than logical items. The subsequent model assessing mentalizing ratings (by-item slope was dropped to achieve convergence) also detected a significant main effect of condition at both contrasts (C1-Log/Men: $\beta = 0.707$, $SE = 0.051$, $t = 13.871$, $p < 0.001$; C2-Coh/Inc: $\beta = -0.411$, $SE = 0.063$, $t = -6.519$, $p < 0.001$), indicating that incoherent items were perceived as needing less mentalizing than all coherent items and mental state items as needing more mentalizing than logical items, as expected. Lastly, the model assessing global reading time (standardized, log-transformed) detected a significant main effect of condition, but only in the contrast between coherent (logical and mental state inferences) vs. incoherent items (C2-Coh/Inc: $\beta = 0.013$, $SE = 0.006$, $t = 2.225$, $p = 0.029$). This effect indicates that incoherent items were generally read more slowly than coherent items, but there was no detectable reading time difference between logical and mental state items. Taken together, we found evidence that our inference type manipulation effectively dissociated the three types of inferences on coherence and mentalizing.

Coherence rating

First, we computed a linear mixed-effects regression model to predict coherence rating (1 = no coherence, 5 = full coherence), where participants were asked whether there was a relationship between the two sentences of the item. This model detected a two-way interaction between the inference type and language group specifically at the second contrast (C2-Coh/Inc: $\beta = 0.095$, $SE = 0.015$, $t = 6.248$, $p < 0.001$). Here, L1 and L2 readers differed in their perceptions of inferences that were generally coherent (mean of logical and mental state) from inferences that were incoherent. Follow-up t-tests on only incoherent vs. logical/mental state item subsets revealed that averages for both subsets differed across L1 and L2 readers (Incoherent: $t(2445.1) = -3.01$, $p = 0.003$; Logical + mental state: $t(5037) = 8.09$, $p < 0.001$). Thus, L2 readers rated incoherent items as more coherent, and they rated all coherent items (mean of logical and mental state inferences) as less coherent, compared to L1 readers. We did not detect any effects of language diversity on coherence ratings. The fixed effects of this model accounted for 72.4% variance of the data (marginal R^2), and the fixed and random effects of this model accounted for 76.9% of variance of the data (conditional R^2).

Mentalizing rating

Next, we computed a linear mixed-effects regression model to predict mentalizing rating (1 = no mentalizing, 5 = full mentalizing), where participants were asked to what extent the relationship between the sentences depended on mentalizing, or the understanding of the character's thoughts, behaviors, goals, and emotions. This model detected several significant interactions between inference type, language diversity, and language group: inference type x language diversity at both contrast levels (C1: Log/Men = $\beta = 0.087$, $SE = 0.031$, $t = 2.827$, $p = 0.005$; C2: Coh/Inc = $\beta = -0.074$, $SE = 0.018$, $t = -4.165$, $p < 0.001$), inference type x language group at both contrast levels (C1: Log/Men = $\beta = -0.173$, $SE = 0.040$, $t = -4.276$, $p < 0.001$; C2: Coh/Inc = $\beta = -0.078$, $SE = 0.023$, $t = -3.332$, $p < 0.001$), and inference type x language diversity x language group only at the second contrast (C2: Coh/Inc = $\beta = 0.114$, $SE = 0.023$, $t = 4.798$, $p < 0.001$). We will first review the separate two-way interactions with language diversity and language group (i.e., no 3-way interaction) that affirmatively involved logical and mental state inferences (C1: Log/Men), given that dissociating these inference types was of primary interest in this paper.

From observing the interaction between inference type and language diversity in Figure 1 (panel B), it seemed that greater language diversity patterned with higher perceptions of mentalizing only to mental state inferences (i.e., not logical inferences), regardless of whether one was a first or second language reader (panel C). To statistically confirm this, we conducted follow-up simple linear regressions on the logical and mental state inference subsets. These models confirmed that language diversity was only related to changes in mentalizing rating of mental state ($\beta = 0.121$, $SE = 0.027$, $t = 4.56$, $p < 0.001$) and not logical inferences ($\beta = 0.034$, $SE = 0.028$, $t = 1.22$, $p = 0.221$). This suggested that greater language diversity enhanced mentalizing ratings of mental state inferences.

We also observed the interaction between inference type and language group from Figure 1 (panel A). Here, the difference between logical and mental state inferences appeared smaller for second language readers than it did for first language readers, regardless of language diversity. To test this observation and better understand the nature of this interaction (i.e., did responses to logical, mental state, or both types of inferences change as a function of language group?), we conducted two follow-up t-tests on logical vs. mental state items. Language group was significant among logical items ($t(2550) = -4.742$, $p < 0.001$), but not mental state items ($t(2625) = -1.659$, $p = 0.097$), indicating that logical, but not mental state, inferences were perceived as higher in mentalizing by L2 readers than L1 readers. Thus, the difference in mentalizing ratings of logical and mental state inferences between L1 and L2 readers was driven by second language readers attributing MORE MENTALIZING TO LOGICAL ITEMS than first language readers. Moreover, a third follow-up t-test indicated that there was no difference in mentalizing ratings to incoherent items between L1 and L2 readers ($t(2625.6) = 1.933$, $p = 0.053$).

Lastly, the model detected a significant three-way interaction between inference type, language diversity, and language group among all coherent (logical and mental state inferences) and incoherent items. Visual inspection of this relationship (Figure 1, panel C) revealed that this interaction may have been driven by responses to incoherent items as a function of individual differences in bilingual language experience. Indeed, recomputing this model within a subset of only logical and mental state inferences (and removing inference type from the interaction) did not

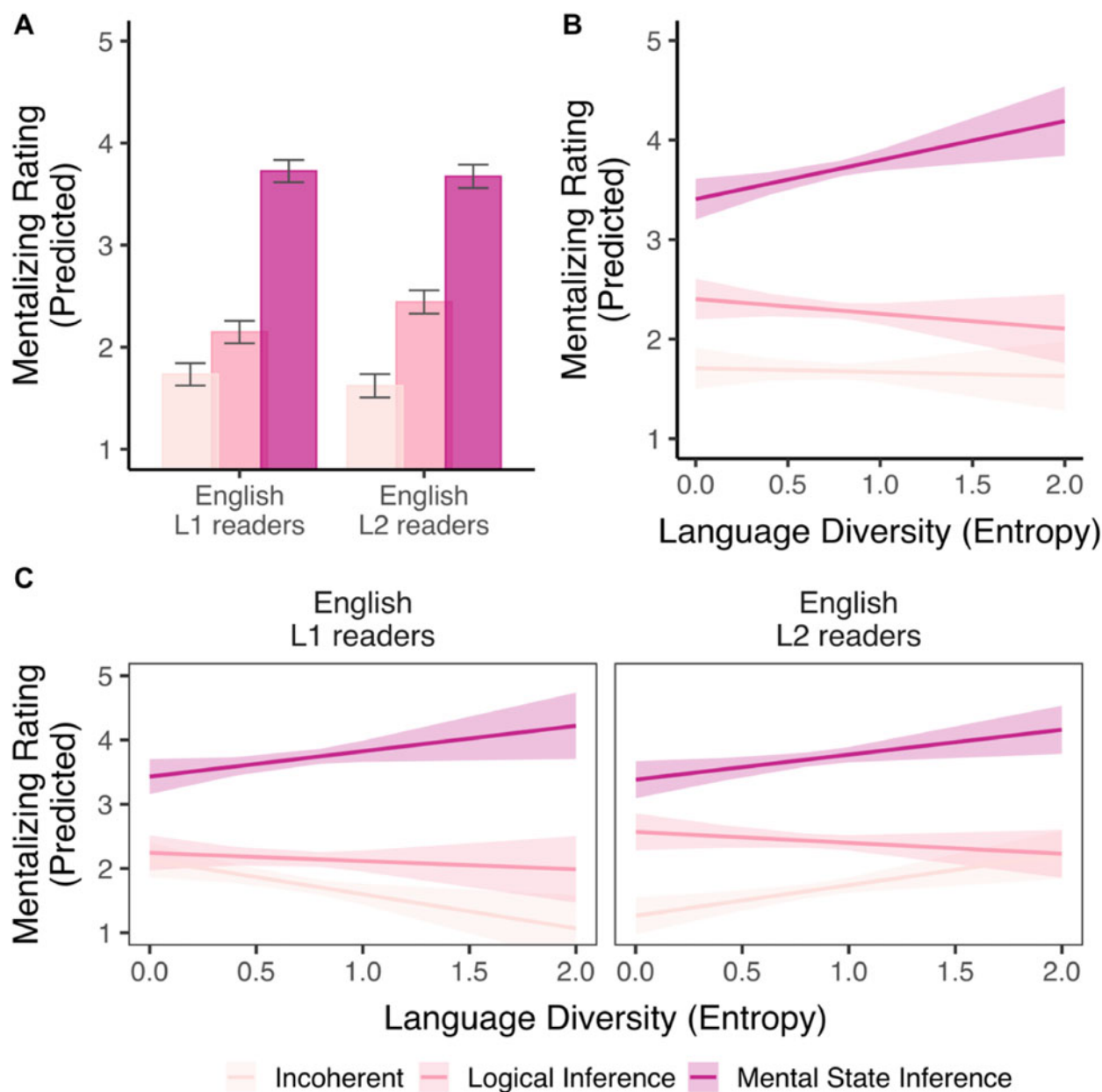


Figure 1. Model-predicted mentalizing rating as a function of (A) language group (B) language diversity (C) language group and language diversity. Error lines represent plus or minus one standard error of the mean. Model syntax: $\text{lmer}(\text{Mentalizing rating} \sim \text{inference type} * \text{scale}(\text{language diversity}) * \text{language group} + \text{scale}(\text{trial order}) + \text{scale}(\text{daily percent English use}) + (1 | \text{subject}) + (1 | \text{item}), \text{df}, \text{contrasts} = \text{list}(\text{inference type} = \text{cHelmert}), \text{REML} = \text{F})$

return a significant interaction between language diversity and language group, thus confirming that any relationship with the individual difference variables stemmed from the incoherent items. The fixed effects of this model accounted for 26.8% variance of the data (marginal R^2), and the fixed and random effects of this model accounted for 37% of variance of the data (conditional R^2).

Global reading time

Lastly, we computed a linear mixed-effects regression to predict global reading time (i.e., the total time it took to read both sentences), which was standardized to account for length differences across items (using two approaches, see Supplementary Materials) and log-transformed for normality. This model did not return any

detectable interactions between inference type and the individual difference variables. The fixed effects of this model accounted for 9.1% variance of the data (marginal R^2), and the fixed and random effects of this model accounted for 41% of variance of the data (conditional R^2).

Discussion

We evaluated the potential relationship between mentalizing and continuous individual differences in bilingual language experience. To accomplish this, we implemented three specific goals. First, we evaluated whether mentalizing judgements of logical vs. mental state inferences varied as a function of individual differences in bilingual language diversity. Second, we examined whether this pattern of results was constrained by first vs. second

reading. Third, we assessed the potential differences in global reading time between logical and mental state inferences as a function of language diversity.

The pattern of results first revealed that language diversity related to mentalizing, as predicted. Specifically, greater language diversity, as measured through language entropy, patterned with greater recognition of mentalizing when a mental state inference was needed to understand a character's behavior from a short, written vignette. There was no impact of language diversity on judgments of linguistic coherence, which may suggest that the role of language diversity pertains more to social content than the demands of linguistic processing. Second, the relationship between language diversity and mentalizing varied as a function of whether the reader was a first or second language English speaker; however, this pattern of results seemed to be driven by mentalizing judgments to incoherent items. Of note here was that L1 vs. L2 reading exerted its own influence (without involving language diversity) on mentalizing judgments of logical vs. mental state inferences. Specifically, L2 readers attributed greater need for mentalizing to LOGICAL INFERENCES, suggesting a tendency to overgeneralize situations in which mentalizing would be useful to understand a character's behavior. Moreover, we found that L1 vs. L2 reading also shaped overall perceptions of coherence, but this was only detected when contrasting all coherent items (logical and mental state inferences) from incoherent items. Third, we did not find reading time differences between logical and mental state inferences, though our manipulation check detected reading time differences between coherent and incoherent items. Global reading time was not mediated by individual differences in bilingual language experience. While reaction times to the coherence and mentalizing judgments were recorded, we reported those findings in Supplementary Materials (Supplementary Materials).¹

Manipulation check

We first assessed core differences in the three inference types (logical, mental state, and incoherent) across all readers to ensure that our items reflected the type of inference that we intended. As expected, incoherent items were rated as less coherent than both logical and mental state inferences. Similarly, mental state inferences were rated as less coherent than logical inferences, suggesting a dissociation between the two types of pragmatic inferences: those rooted in thinking about other minds were found to be linguistically less coherent than those involving non-social information.

Next, we compared the extent to which each inference type was recognized as needing mentalizing for comprehension. Here, we found that incoherent items were rated lower in mentalizing than both logical and mental state inferences. Critically, mental state inferences were rated as higher in mentalizing than logical inferences, indicating that the two types of pragmatic inferences could further be dissociated in their degree of reliance on thinking about other minds. Taken together, this manipulation check affirmatively demonstrated that although logical inferences inherently relied on some level of social inferencing (as a result of including a character to maintain parallel structure within an item set as possible), a diverse set of raters were still able to

distinguish between logical and mental state inferences through their mentalizing ratings.

Lastly, we found that global reading time of incoherent items was longer than that for all coherent items, suggesting greater processing demands for short vignettes that did not make sense. Surprisingly, we did not detect any global reading time differences between logical and mental state inferences. However, it is possible that our coarse measurement of reading time across two full sentences may not have been sensitive enough to capture the nuances between these two inference types. Ongoing work using more granular methods, such as eye tracking, will better assess potential reading time differences between logical and mental state pragmatic inferences.

Individual differences in bilingual language experience and mentalizing

The primary question that this work aimed to address is whether continuous individual differences in language diversity, or the extent to which multiple languages are used regularly, related to recognition of mentalizing. Prior work on self-reported language frequency and number of languages known demonstrated a positive association between greater bilingual experience and stronger cognitive empathy or mentalizing abilities (Dewaele & Wei, 2012; Mepham & Martinovic, 2018). Thus, we expected that greater language diversity would also be associated with more accurate mentalizing during this reading task, which is what we found. Language diversity did not co-vary with ratings of linguistic coherence, but – as expected – greater language diversity patterned with higher mentalizing ratings of only mental state inferences. Mentalizing ratings to logical inferences, similarly, did not change with language diversity. Together, these results support the notion that greater language diversity, an index of regularly using multiple languages and potentially encountering novel social situations, selectively related to mentalizing when it was needed to understand the actions of a character.

In forming this interpretation, we considered the possibility that language diversity could have related to the greater inferencing demands of the mental state inferences, as opposed to the social content. For example, relevant research from the discourse processing literature illustrated that intermediately causally related sentences (rated low in relatedness, or “far” inferences) are more difficult than highly causally related sentences (rated high in relatedness, or “near” inferences) to process and remember (e.g., Myers, Shinjo & Duffy, 1987). While the linguistic distance of mental state inferences may have been further than the linguistic distance of logical inferences, our design accounted for this possibility by having participants evaluate all inferences on both mentalizing and linguistic coherence, which was operationally defined as the relatedness between the two sentences. To the extent that this definition conceptually maps onto causal relatedness from Myers et al. (1987) and given that language diversity did not predict linguistic coherence in this study, it is likely that the social content of the mental state inference, as opposed to the linguistic distance of the inference, is at the core of the relationship with bilingual experience. Furthermore, to statistically account for linguistic distance of the inference, we added coherence ratings to our mentalizing individual differences model as a covariate; however, doing so did not alter our pattern of results. Despite the convergent evidence against a linguistic processing account, we encourage future research on mentalizing using

¹Neither coherence nor mentalizing reaction time models detected significant relationships between logical and mental state inferences with individual differences in bilingual language experience.

linguistic stimuli to more systematically consider such demands and how they may impact processing.

For example, given that mentalizing inherently relies on inference-making, and past work has shown that some L2 readers struggle with making complex inferences (e.g., Foucart, Romero-Rivas, Lottie & Costa, 2016), we examined whether the relationship between mentalizing and bilingualism would change depending on whether readers were reading in their first or second language. We detected a significant three-way interaction between inference type, language diversity, and language group, but only when contrasting all coherent (logical and mental state inference) to incoherent items. Follow-up tests revealed that this pattern was driven by responses to incoherent items, which may have reflected differences in general tolerance or openness between first and second language readers depending on their language diversity (e.g., Dewaele & Li, 2013). However, our primary interest was in discerning logical and mental state inferences, and we encourage future research to better understand the nature of this unexpected result.

Interestingly, we observed a relationship between inference type and language group, independent of language diversity, on mentalizing judgments. Here, L2 readers attributed more mentalizing to logical inferences than L1 readers. This result was interesting because it suggested that L2 readers may be over-mentalizing in cases where mentalizing is not necessarily needed for the inference (or may not be the main source of the inference), whereas mentalizing for mental state inferences did not change. Past work from clinical samples has referred to the over-attribution of mentalizing as hypermentalizing and has linked this behavior to excessive attention to external cues stemmed in feelings of vulnerability (Bateman & Fonagy, 2019). Whereas the present sample did not report a history of clinical disorders, it is possible that the experience of reading or operating in a second language may also trigger additional attention to external cues if that language induces vulnerability in any way. For example, the English framing of the task (instructions, conversations with experimenters, language of the text) may have prompted some L2 English readers to implicitly begin seeking additional cues in the text to help them understand, similar to past findings that nonverbal cues like gesture compensate for low proficiency L2 production (Gullberg, 1998). Additionally, there was no difference in L1 and L2 mentalizing ratings to incoherent items. Thus, it was not the case that L2 readers simply rated all items higher in mentalizing, regardless of coherence. Instead, they over-mentalized only among coherent items that were based on logical inferences.

Taken together, we found evidence supporting our prediction that greater language diversity relates to better recognition of mentalizing, specifically for mental state inferences. Additionally, we presented additional evidence of L2 readers over-mentalizing for logical inferences, that we interpret to mean heightened sensitivity to mentalizing among L2 readers, potentially arising from the L2 context demanding more mentalizing (e.g., Foucart et al., 2015). However, we also observed that L2 readers demonstrated less dissociation between coherent and incoherent items compared to L1 readers who exhibited crisper separation — though there was no detectable group difference in coherence between logical and mental state inferences. Thus, it is possible that L2 readers may indeed struggle with the linguistic demands of finding coherence in a set of sentences, but perhaps their over-application of mentalizing compensates to aid in their overall understanding of the situation at hand.

Potential mechanisms

Collectively, these results suggest that readers who regularly use more languages (experience greater language diversity) also demonstrate more flexible social cognition, as measured by greater recognition of mentalizing. As discussed in other work (e.g., Antoniou, 2019; Schroeder, 2018; Tiv et al., 2019b, 2020), there are a number of potential mechanisms underlying the relationship between bilingualism and mentalizing. These include (1) greater metalinguistic awareness, including insight on the flexible nature of language and the many names that single concepts can have across languages (e.g., Goetz, 2003) (2) enhanced executive functions, such that greater cognitive control aids in downregulating one's own mental state and adopting another's mental state (e.g., Goetz, 2003; Kovács, 2009), and (3) strengthened social-pragmatic flexibility, based on an understanding that people themselves can speak different languages, come from different backgrounds, and consequently can think differently (e.g., Fan et al., 2015). We now briefly discuss how our data relate to these frameworks.

Some claim that strong mentalizing capacities are born out of greater linguistic proficiency, whether that be in a single language or across multiple languages, which cultivates insight of language on a conceptual level (e.g., Goetz, 2003; Milligan et al., 2007; Pyers & Senghas, 2009; Rutherford, Wareham, Vrouva, Mayes, Fonagy & Potenza, 2012; Warnell & Redcay, 2019). However, in contrast to this metalinguistic awareness account, all of our regression models coded for daily English usage (the language of the task), and they still detected a significant relationship between language diversity and mentalizing. We interpret this to mean that the amount of time one spends in a given language does not discount the overall experience of juggling multiple languages, as measured by language diversity. In other words, it seems likely that some general aspect of language diversity, such as probabilistically casting predictions in different contexts or centrally managing language use across contexts, drives this relationship. This may support the executive control account, based on the cognitive implications of language switching, or it may support the social-pragmatic flexibility account, related to interacting with a more diverse set of people and languages.

Past research on bilingual cognitive advantages has been contentious (reviewed in Baum & Titone, 2014). Indeed, Tiv et al. (2020) detected a significant relationship between bilingualism and irony comprehension despite statistically controlling for a core component of the central executive functions. In contrast, results from other studies on the positive SOCIAL effects of bilingualism, appear more consistent (e.g., Dewaele & Wei, 2012; Mephram & Martinovic, 2018; Ramírez-Esparza et al., 2020). For example, Ikizer and Ramírez-Esparza (2018) tested whether U.S. and Canadian bilingual adults' social interactions were mediated by increased social flexibility, or the ability to switch with ease and adapt between different social environments, as compared to U.S. and Canadian monolingual adults. The results showed that bilinguals self-reported greater social flexibility, compared to monolinguals, and this difference mediated the frequency of their subsequent social interactions. This evidence suggests that having more opportunities to engage in novel social situations through the regular exercising of multiple languages, or greater language diversity, may expand one's understanding of the world and in turn their mentalizing capacity (reviewed in Ramírez-Esparza et al., 2020). Indeed, even passive exposure to greater language diversity in the environment promotes

mentalizing capacities among functionally monolingual children raised in multilingual areas (Fan *et al.*, 2015).

The social-pragmatic flexibility account is also consistent with social psychological research demonstrating that identification of self-other differences (i.e., parsing one's own mental state from another's) is a critical component of successful mentalizing (Decety & Sommerville, 2003; Higgins, 1981; Mitchell, 2009; Tamir & Mitchell, 2010). Todd and colleagues (2011; 2016) contend that recognizing self-other similarities in mental state may be a crucial first step in mentalizing, but in order to formulate accurate inferences, differences in mental state must also be computed. In other words, successful mentalizing involves a comparative process to analyze the similarities and differences between the perceiver and the one being perceived (Todd *et al.*, 2011, 2016). Bilinguals who experience greater language diversity may be also experience other forms of diversity in their daily lives (e.g., racial, gender, class, nationality). Encoding these environmental differences may cue high language diversity readers to exercise greater mentalizing than low language diversity readers. Our follow-up work is actively pursuing these potentially social mechanisms that may link bilingual sociolinguistic diversity and mentalizing.

Limitations and strengths

We acknowledge some limitations of this work. First, while condition was within-subject, we also utilized a between-subjects design, which did not allow us to compare L1 vs. L2 reading within the same individual. As a result, there might have been confounding differences between our two groups that contributed to the results. We strove to mitigate this possibility by including daily percent English use in our statistical models to control for usage of the language of the task. Second, we did not explicitly probe comprehension of the items; instead, we used coherence as an indicator of whether readers understood the item. The rationale was that if the correct inference is made (i.e., what the experimenters intended), then that would be reflected in the coherence score. Nevertheless, it is possible that readers made a different inference from what the experimenters intended, which could give rise to a weaker or stronger coherence; however, we intentionally designed the materials to feature simple, high frequency language so that they would be understandable by our sample of proficient bilinguals. Third, given that our primary outcome variables were the rating judgments, it is possible that performance on these measures reflected later processing as opposed to in-the-moment mentalizing. Still, the results converge with past findings from our group demonstrating a positive relationship between continuous assessments of bilingual language experience and mentalizing on on-line comprehension tasks (Tiv *et al.*, 2020).

These limitations notwithstanding, we highlight several strengths of the present study. First, we bridged traditions and methods from social cognition and the language sciences (bilingualism, discourse processes) to address inference making as a simultaneously linguistically and socially rooted process. In merging these lines of inquiry, we presented a task design that was familiar to most participants: reading short vignettes. Whether it be reading text messages, emails, articles, books, or city signs, most literate individuals read in some capacity on a daily basis. Thus, by targeting mental state inference-making through reading, we capitalized on a natural and ecologically valid social process that is otherwise missing in many traditional assessments of

mentalizing, perspective-taking, or Theory of Mind (e.g., Bloom & German, 2000; Ferguson, Apperly, Ahmad, Bindemann & Cane, 2015). Ongoing work from our group is utilizing eye tracking to measure on-line inference making to better discern the time course of logical and mental state inferences.

Additionally, as discussed, our study did not include a monolingual group, which was decided for several reasons. First, despite visibility in the United States, monolinguals make up a small proportion of the world's language users (Grosjean, 2010), which does not warrant the tethering of results back to a default monolingual "control" group, or monolingual hegemony (Ortega, 2018; Tiv, Kutlu & Titone, 2021; Vaid & Meuter, 2016, 2017). Second, monolinguals and bilinguals often have many differences between them beyond the simple number of languages known (e.g., culture, education, SES), which could serve as confounds in their task performance. Altogether, instead of comparing our bilinguals to a monolingual sample, we capitalized on a continuous assessment. Through this, we found that more diverse bilingual language experience co-varied with stronger mentalizing scores, much like past work that has revealed bilingual children and adults outperformed monolinguals on mentalizing and Theory of Mind tasks (Navarro & Conway, 2020; Rubio-Fernández & Glucksberg, 2012; Schroeder, 2018).

Conclusion

To conclude, the present work established a link between bilingual language experience and mentalizing. Our main results indicated greater language diversity patterned with greater mentalizing judgments of mental state inferences. We also found that, whereas L2 readers displayed less crisp linguistic coherence dissociations between coherent and incoherent inferences, they rated logical inferences as higher in mentalizing than L1 readers, potentially compensating for any linguistic challenges with greater reliance on social information.

These results contribute to a growing body of work that aims to highlight the positive social benefits of bilingualism (see Ramírez-Esparza *et al.*, 2020), such as mentalizing and social cognition. Mentalizing humanizes people – it breaks us out of our own minds and brings us into the minds of others. Though it has been historically understudied in the field of bilingualism, mentalizing may provide a mechanism to explain why bilinguals demonstrate attenuated other-race effects (Burns, Tree, Chan & Xu, 2019) and less racial bias (Singh, Quinn, Qian & Lee, 2020). These nascent areas of interdisciplinary investigation are promising not only for their scientific novelty, but also for their potential real-world implications, such as mitigating intergroup conflict.

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728921000225>

Acknowledgements. This research was funded by the Natural Sciences and Engineering Research Council of Canada (Grant 261769-13), the Fonds de Recherche du Québec: Société et Culture, and the Canada Research Chairs Program. The authors would like to thank Jasper Evans, Ruo Feng, Jason Gullifer, Mikkel Kranker Jørgensen, Mehrshid Kiazand, Ethan Kutlu, Pauline Palma, Christina Rigas, Charlotte Rossi-McCunn, Mehrnaz Tiv, Mehran Tiv, and Naomi Vingron for their thoughtful feedback in developing the stimuli and this manuscript.

References

- Antoniou K** (2019) Multilingual pragmatics: Implicature comprehension in adult L2 learners and multilingual children. In Taguchi N (ed), *The Routledge Handbook of Pragmatics and Second Language Acquisition*. New York: Routledge, pp. 495–510.
- Antoniou K and Katsos N** (2017) The effect of childhood multilingualism and bilingualism on implicature understanding. *Applied Psycholinguistics* **38**, 787–833. <https://doi.org/10.1017/S014271641600045X>
- Apperly IA** (2008) Beyond Simulation-Theory and Theory-Theory: Why social cognitive neuroscience should use its own concepts to study “theory of mind.” *Cognition* **107**, 266–283. <https://doi.org/10.1016/j.cognition.2007.07.019>
- Astington JW and Gopnik A** (1988) Knowing you’ve changed your mind: Children’s understanding of representational change. In Astington JW, Harris P and Olson D (eds), *Developing theories of mind*. Cambridge University Press, pp. 193–206.
- Baimel A, Birch SA and Norenzayan A** (2018) Coordinating bodies and minds: Behavioral synchrony fosters mentalizing. *Journal of Experimental Social Psychology* **74**, 281–290. <https://doi.org/10.1016/j.jesp.2017.10.008>
- Banasik N** (2013) Non-literal speech comprehension in preschool children - an example from a study on verbal irony. *Psychology of Language and Communication* **17**, 309–323. <https://doi.org/10.2478/plc-2013-0020>
- Baron-Cohen S, Leslie AM and Frith U** (1985) Does the autistic child have a “theory of mind”? *Cognition* **21**(1), 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Baron-Cohen S, Wheelwright S, Hill J, Raste Y and Plumb I** (2001) The “reading the mind in the eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology Psychiatry* **42**, 241–251. <https://doi.org/10.1017/S0021963001006643>
- Baron-Cohen S and Wheelwright S** (2004) The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders* **34**, 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Barr DJ, Levy R, Scheepers C and Tily HJ** (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* **68**, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bateman A and Fonagy P** (2019) Mentalization-based treatment for borderline and antisocial personality disorder. In Kealy D and Ogronczuk JS (eds), *Contemporary Psychodynamic Psychotherapy: Evolving Clinical Practice*. Academic Press, pp 133–148.
- Bates D, Maechler M, Bolker B and Walker S** (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7.
- Baum S and Titone D** (2014) Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics* **35**, 857–894. doi:10.1017/s0142716414000174
- Bice K and Kroll JF** (2019) English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language* **196**, 104644. doi:10.1016/j.bandl.2019.104644
- Bloom P and German TP** (2000) Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* **77**, 25–31. [https://doi.org/10.1016/S0010-0277\(00\)00096-2](https://doi.org/10.1016/S0010-0277(00)00096-2)
- Burns EJ, Tree J, Chan AH and Xu H** (2019) Bilingualism shapes the other race effect. *Vision Research* **157**, 192–201.
- Conway JR, Coll MP, Cuve HC, Koletsi S, Bronitt N, Catmur C and Bird G** (2019) Understanding how minds vary relates to skill in inferring mental states, personality, and intelligence. *Journal of Experimental Psychology: General* **149**, 1032–1047. <https://doi.org/10.1037/xge0000704>
- Davis MH** (1980) A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology* **10**, 85
- De Bruin A** (2019) Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences* **9**, 33. <https://doi.org/10.3390/bs9030033>
- Decety J and Sommerville JA** (2003) Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences* **7**, 527–533. <https://doi.org/10.1016/j.tics.2003.10.004>
- Dewaele JM and Wei L** (2012) Multilingualism, empathy and multicompetence. *International Journal of Multilingualism* **9**, 352–366. <https://doi.org/10.1080/14790718.2012.714380>
- Dewaele JM and Li W** (2013) Is multilingualism linked to a higher tolerance of ambiguity?. *Bilingualism: Language and Cognition* **16**, 231–240. doi:10.1017/S1366728912000570
- Dumontheil I, Apperly IA and Blakemore SJ** (2010) Online usage of theory of mind continues to develop in late adolescence. *Developmental Science* **13**, 331–338. <https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Dziobek I, Rogers K, Fleck S, Bahnemann M, Heekeren HR, Wolf OT and Convit A** (2008) Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *Journal of Autism Developmental Disorders* **38**, 464–73.
- Fan SP, Liberman Z, Keysar B and Kinzler KD** (2015) The Exposure Advantage: Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological Science* **26**, 1090–1097. <https://doi.org/10.1177/0956797615574699>
- Ferguson HJ, Apperly I, Ahmad J, Bindemann M and Cane J** (2015) Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition* **139**, 50–70. <https://doi.org/10.1016/j.cognition.2015.02.010>
- Ferstl EC and Von Cramon DY** (2002) What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *NeuroImage* **17**, 1599–1612. <https://doi.org/10.1006/nimg.2002.1247>
- Filippova E and Astington JW** (2010) Children’s understanding of social-cognitive and social-communicative aspects of Discourse Irony. *Child Development* **81**, 913–928. <https://doi.org/10.1111/j.1467-8624.2010.01442.x>
- Fletcher P, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RSJ, and Frith CD** (1995) Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. *Cognition* **57**, 109–128.
- Foucart A, Garcia X, Ayguasanosa M, Thierry G, Martin C and Costa A** (2015) Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia* **75**, 291–303. <https://doi.org/10.1016/j.cogdev.2016.10.003>
- Foucart A, Romero-Rivas C, Lottie B and Costa A** (2016) Brain & Language Discourse comprehension in L2 : Making sense of what is not explicitly said. *Brain and Language* **163**, 32–41. <https://doi.org/10.1016/j.bandl.2016.09.001>
- Frith CD and Frith U** (1999) Interacting minds—a biological basis. *Science* **286**, 1692–1695. doi:10.1126/science.286.5445.1692
- Gaudreau G, Monetta L, Maccoir J, Poulin S, Laforce RJ and Hudon C** (2015) Mental State Inferences Abilities Contribution to Verbal Irony Comprehension in Older Adults with Mild Cognitive Impairment. *Behavioural Neurology*, 1–9. <https://doi.org/10.1155/2015/685613>
- Goetz PJ** (2003) The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition* **6**, 1–15. <https://doi.org/10.1017/S1366728903001007>
- Graesser AC, Singer M and Trabasso T** (1994) Constructing inferences during narrative text comprehension. *Psychological review* **101**, 371. <https://doi.org/10.1037/0033-295X.101.3.371>
- Grosjean J** (2010) Attitudes and Feelings about Bilingualism. In Grosjean J (ed), *Bilingual: Life and Reality*. Harvard University Press, pp 97–107.
- Gullberg M** (1998) *Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish* (Vol. 35). Lund University.
- Gullifer JW and Titone D** (2018) Compute language entropy with {languageEntropy}. Retrieved from <https://github.com/jasongullifer/languageEntropy>. doi:10.5281/zenodo.1406201
- Gullifer JW and Titone D** (2019) Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition* **23**, 283–294. <https://doi.org/10.1017/S1366728919000026>
- Gullifer JW and Titone D** (2020) Engaging proactive control: Influences of diverse language experiences using insights from machine learning. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000933>
- Harris LT** (2017) *Invisible mind: flexible social cognition and dehumanization*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262035965.001.0001
- Harris RJ and Monaco GE** (1978) Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General* **107**, 1–22. <https://doi.org/10.1037/0096-3445.107.1.1>
- Higgins ET** (1981) Role taking and social judgment: Alternative developmental perspectives and processes. In Flavell R and Parke R (eds), *Social Cognitive Development: Frontiers and Possible Futures*. Cambridge University Press, pp 119–153.

- Horiba Y** (1996) Comprehension Processes in L2 Reading. *Studies in Second Language Acquisition* **18**, 433–473. <https://doi.org/10.1177/0145482990300906>
- Hsin L and Snow C** (2017) Social perspective taking: A benefit of bilingualism in academic writing. *Reading and Writing* **30**, 1193–1214. <https://doi.org/10.1007/s11145-016-9718-9>
- Ikizer EG and Ramírez-Esparza N** (2018) Bilinguals' social flexibility. *Bilingualism: Language and Cognition* **21**, 957–969. <https://doi.org/10.1017/S1366728917000414>
- Johnson J and Rosano T** (1993) Relation of cognitive style to metaphor interpretation and second language proficiency. *Applied psycholinguistics* **14**, 159–175. <https://doi.org/10.1017/S014271640000953X>
- Johnson-Laird PN** (1993) The mental representation of the meaning of words. In Goldman AI (ed), *Readings in Philosophy and Cognitive Science*. MIT Press.
- Jolliffe D and Farrington DP** (2006) Development and validation of the Basic Empathy Scale. *Journal of Adolescence* **29**, 589–611. <https://doi.org/10.1016/j.adolescence.2005.08.010>
- Kanske P, Böckler A, Trautwein FM, Parianen Lesemann FH and Singer T** (2016) Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective neuroscience* **11**, 1383–1392. <https://doi.org/10.1093/scan/nsw052>
- Kidd DC and Castano E** (2013) Reading literary fiction improves theory of mind. *Science* **342**, 377–380. doi: 10.1126/science.1239918
- Kispaal A** (2008) Effective teaching of inference skills for reading. Retrieved from: <https://files.eric.ed.gov/fulltext/ED501868.pdf>
- Kovács Á. M.** (2009) Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science* **12**, 48–54. <https://doi.org/10.1111/j.1467-7687.2008.00742.x>
- Lavoie MA, Vistoli D, Sutliff S, Jackson PL and Achim AM** (2016) Social representations and contextual adjustments as two distinct components of the Theory of Mind brain network: evidence from the REMICS task. *Cortex* **81**, 176–191. <https://doi.org/10.1016/j.cortex.2016.04.017>
- Matuschek H, Kliegl R, Vasishth S, Baayen H and Bates D** (2017) Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* **94**, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Mepham KD and Martinovic B** (2018) Multilingualism and out-group acceptance: The mediating roles of cognitive flexibility and deprovincialization. *Journal of Language and Social Psychology* **37**, 51–73. <https://doi.org/10.1177/0261927X17706944>
- Milligan K, Astington JW and Dack LA** (2007) The Language Demographics of Amazon Mechanical Turk. *Child Development* **78**, 622–646. https://doi.org/10.1162/tacl_a_00167
- Mitchell JP** (2009) Social psychology as a natural kind. *Trends in Cognitive Sciences* **13**, 246–251. <https://doi.org/10.1016/j.tics.2009.03.008>
- Myers JL, Shinjo M and Duffy SA** (1987) Degree of causal relatedness and memory. *Journal of Memory and Language* **26**, 453–465. [https://doi.org/10.1016/0749-596X\(87\)90101-X](https://doi.org/10.1016/0749-596X(87)90101-X)
- Nadig A and Ozonoff S** (2007) Bridging inferences in high-functioning autism: A reading time study comparing mental and physical inferences. Poster presented at the 32nd Annual Boston University Conference on Language Development, Boston, MA, USA.
- Navarro E and Conway ARA** (2020) Adult Bilinguals Outperform Monolinguals in Theory of Mind. *PsyArXiv*. <https://doi.org/10.31234/osf.io/nyt3w>
- Ortega L** (2018) Ontologies of language, Second Language Acquisition, and world Englishes. *World Englishes* **37**, 64–79. <https://doi.org/10.1111/weng.12303>
- Pino MC and Mazza M** (2016) The use of “literary fiction” to promote mentalizing ability. *PLoS One* **11**, <https://doi.org/10.1371/journal.pone.0160254>
- Poynor DV and Morris RK** (2003) Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 3. <https://doi.org/10.1037/0278-7393.29.1.3>
- Premack D and Woodruff G** (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* **1**, 515–526. <https://doi.org/10.1016/j.celrep.2011.1011.1001.7>
- Pyers JE and Senghas A** (2009) Language Promotes False-Belief Understanding: Evidence from Learners of a New Sign Language. *Psychological Science* **20**, 805–812. <https://doi.org/10.1111/j.1467-9280.2009.02377.x>
- Rai MK, Loschky LC, Harris RJ, Peck NR and Cook LG** (2011) Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning* **61**, 187–218. <https://doi.org/10.1111/j.1467-9922.2010.00592.x>
- Ramírez-Esparza N, García-Sierra A and Jiang S** (2020) The current standing of bilingualism in today's globalized world: A socio-ecological perspective. *Current Opinion in Psychology* **32**, 124–128. <https://doi.org/10.1016/j.copsyc.2019.06.038>
- Rubio-Fernández P and Glucksberg S** (2012) Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning Memory and Cognition* **38**, 211–217. <https://doi.org/10.1037/a0025162>
- Rutherford HJ, Wareham JD, Vrouva I, Mayes LC, Fonagy P and Potenza MN** (2012) Sex differences moderate the relationship between adolescent language and mentalization. *Personality Disorders: Theory, Research, and Treatment* **3**, 393. <https://doi.org/10.1037/a0028938>
- Sarfati Y, Hardy-Baylé MC, Besche C and Widlöcher D** (1997) Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophrenia Research*. **25**, 199–209.
- Saxe R and Kanwisher N** (2003) People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Neuroimage* **19**, 1835–1842. [https://doi.org/10.1016/S1053-8119\(03\)00230-1](https://doi.org/10.1016/S1053-8119(03)00230-1)
- Schroeder SR** (2018) Do Bilinguals Have an Advantage in Theory of Mind? A Meta-Analysis. *Frontiers in Communication* **3**, 1–8. <https://doi.org/10.3389/fcomm.2018.00036>
- Singh L, Quinn PC, Qian M and Lee K** (2020) Bilingualism is associated with less racial bias in preschool children. *Developmental Psychology* **56**, 888–896. <https://doi.org/10.1037/dev0000905>
- Slaughter V and Repacholi B** (2003) Individual differences in theory of mind: What are we investigating?. *Individual differences in theory of mind: Implications for typical and atypical development*, 1–13. <https://doi.org/10.4324/9780203488508>
- Snow C** (2002) *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. The RAND Corporation. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1465.html
- Sperber D and Wilson D** (2002) Pragmatics, modularity and mind-reading. *Mind & Language* **17**, 3–23. <https://doi.org/10.1111/1468-0017.00186>
- Sundaray S, Marinis T and Bose A** (2018) Comprehending non-literal language: effects of aging and bilingualism. *Frontiers in Psychology* **9**, 2230. <https://doi.org/10.3389/fpsyg.2018.02230>
- Tamir DI and Mitchell JP** (2010) Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences* **107**, 10827–10832. <https://doi.org/10.1073/pnas.1003242107>
- Tiv M, Kutlu E and Titone D** (2021) Bilingualism Moves us Beyond the Ideal Speaker Narrative in Cognitive Psychology. In Francis, W.S. (Ed.). *Bilingualism Across the Lifespan: Opportunities and Challenges for Cognitive Research in a Global Society* (1st ed.). Routledge. <https://doi.org/10.4324/9781315143996>
- Tiv M, Deodato F, Rouillard V, Wiebe S and Titone D** (2020) Second language experience impacts first language irony comprehension among bilingual adults. *PsyArxiv*. <https://doi.org/10.31234/osf.io/3wu9r>
- Tiv M, Gonnerman L, Whitford V, Friesen D, Jared D and Titone D** (2019a) Figuring out how verb-particle constructions are understood during L1 and L2 reading. *Frontiers in psychology* **10**, 1733. <https://doi.org/10.3389/fpsyg.2019.01733>
- Tiv M, Rouillard V, Vingron N, Wiebe S and Titone D** (2019b) Global Second Language Proficiency Predicts Self-Perceptions of General Sarcasm Use Among Bilingual Adults. *Journal of Language and Social Psychology* **38**, 459–478. <https://doi.org/10.1177/0261927X19865764>
- Todd AR, Hanko K, Galinsky AD and Mussweiler T** (2011) When focusing on differences leads to similar perspectives. *Psychological Science* **22**, 134–141. <https://doi.org/10.1177/0956797610392929>
- Todd AR, Simpson AJ and Tamir DI** (2016) Active perspective taking induces flexible use of self-knowledge during social inference. *Journal of Experimental Psychology: General* **145**, 1583–1588. <https://doi.org/10.1037/xge0000237>
- Vaid J and Meuter R** (2016) Not through a glass darkly: Refocusing the psycholinguistic study of bilingualism through a ‘bivocal’ lens. In Cook V and Wei L (eds.), *The Cambridge handbook of linguistic multi-competence*. Cambridge University Press, pp 77–96.

- Vaid J and Meuter RF** (2017) Languages without borders. In Libben M, Goral M and Libben G (eds), *Bilingualism: A Framework for Understanding the Mental Lexicon*. John Benjamins, pp 7–26.
- Valle A, Massaro D, Castelli I, Sangiuliano Intra F, Lombardi E, Bracaglia E and Marchetti A** (2016) Promoting mentalizing in pupils by acting on teachers: Preliminary Italian evidence of the “Thought in Mind” Project. *Frontiers in Psychology* 7, 1213. <https://doi.org/10.3389/fpsyg.2016.01213>
- Van Overwalle F** (2011) A dissociation between social mentalizing and general reasoning. *Neuroimage* 54, 1589–1599. <https://doi.org/10.1016/j.neuroimage.2010.09.043>
- Verhagen J, Grassmann S and Küntay AC** (2017) Monolingual and bilingual children’s resolution of referential conflicts: Effects of bilingualism and relative language proficiency. *Cognitive Development* 41, 10–18. <https://doi.org/10.1016/j.cogdev.2016.10.003>
- Warnell KR and Redcay E** (2019) Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition* 191, 103997. <https://doi.org/10.1016/j.cognition.2019.06.009>
- Zwaan RA and Brown CM** (1996) The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes* 21, 289–327. <https://doi.org/10.1080/01638539609544960>